Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 175 (2020) 331–339

Procedia Computer Science

www.elsevier.com/locate/procedia

The 15th International Conference on Future Networks and Communications (FNC)
August 9-12, 2020, Leuven, Belgium

Towards Privacy-anomaly Detection: Discovering Correlation between Privacy and Security-anomalies

Muhammad Imran Khan[a,*], Simon N. Foley[b], Barry O'Sullivan[a]

[a]*Insight Centre for Data Analytics, School of Computer Science and Information Technology, University College Cork, Ireland.*
[b]*Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Gjøvik, Norway.*

Abstract

In this paper a notion of *privacy-anomaly detection* is presented where *normative privacy* is modelled using *k*-anonymity. Based on the model, normative privacy-profiles are constructed, and deviation from normative privacy-profile at run-time is labelled as a *privacy-anomaly*. Furthermore, the paper investigates whether there is a correlation between security-anomalies and privacy-anomalies, that is, whether the privacy-anomalies labelled by privacy-anomaly detection system are detected by conventional security-anomaly detection system used for detecting malicious accesses to databases by insiders.

Keywords: Electronic Privacy; Anonymization; Anomaly Detection; *k*-anonymity; Relational Databases;

1. Introduction

One of the challenges in releasing data for analytic is of safeguarding the privacy of individuals whose data is being released. Privacy preservation is achieved using data anonymization. However, we have seen numerous incidents where the privacy was compromised due to poor anonymization of released data, for example, the popular case of Netflix [19], AOL [2] and de-anonymization of NYC taxi data [8]. For this reason researchers have devised privacy definitions, such as *k*-anonymity [23], *l*-diversity [17], *t-closeness* [16], and differential privacy [7], to provide formal guarantees.

We are interested in an anomaly-detection based approach to privacy. Anomaly detection techniques have been widely used in many domains, such as networks and Database Management Systems (DBMS) security to detect attacks [21, 9, 10, 13]. In principle, anomaly detection techniques have the potential to detect

* Corresponding author. Tel.: +353-(0)21-420-5978 .
E-mail address: imran.khan@insight-centre.org

1877-0509
10.1016/j.procs.2020.07.048

zero-day or unknown attacks [20]. Anomaly detection techniques work by looking for a deviation from normative behaviour. Thus, at the heart of an anomaly detection system is a model of normative behaviour. In literature, attempts have been made to capture different aspects of normative behaviours to have an accurate model. We put forward a unique perspective where we model normative from privacy perspective.

In this paper, we propose the notion of *privacy-anomaly detection* that is based on extracting parameters of privacy definitions from logs of past behaviour and building privacy-profiles. Additionally, in this paper, we introduce the notion of *privacy-anomaly* that is a deviation from the profile constructed by extracting parameters of privacy definitions. The main idea is that we analyze past behaviour, which gives us a model that can be used to check subsequent behaviours, where past behaviour is the interaction between the database and the users querying that database. The paper also considers the question of whether there is a correlation between security-anomalies and privacy-anomalies. For ease of exposition, this paper uses k-anonymity to formulate a definition of normative privacy. The scenario that we present in this paper is to model the value for k, in particular, we look at the output of past queries and based on those output tables, we infer the value of k.

The paper is organized as follows. Section 2 provides relevant background on anomaly detection and privacy models. Section 3 defines the notion of a privacy-anomaly detection along with its naïve instantiation based on k-anonymity. Section 4 explores whether there is a correlation between security-anomalies and privacy-anomalies. Conclusions are drawn in Section 5.

2. Background

This section covers the background required for the proposed privacy-anomaly detection model by describing an abstract design of anomaly detection techniques and discusses the adopted privacy model for this work.

2.1. Anomaly Detection

Anomaly detection techniques have been widely used in saveral domains; for example, the popular application domains include intrusion detection, image processing, sensor networks, medical anomaly detection and fault detection [4]. Intrusion detection can be further divided into sub-domains depending on the context in which the intrusion detection systems is deployed for instance database intrusion detection systems where anomaly detection techniques are used to detect malicious accesses to Database Management Systems (DBMS) [10, 21]. Another example is of network intrusion detection systems where anomaly detection techniques monitor network traffic to detect attacks [3, 29]. Typically anomaly detection technique has two phases, that is a training phase (learning phase) and a detection phase. In training (learning phase), a normative profile of normative behaviour is built. In the detection phase, 'activities' are checked if they belong to the normative profile if they deviate then this deviation is labelled as an anomaly. The challenge in anomaly detection is of accurately modelling behaviour as it is possible that one can capture some aspects of normative behaviour and misses some aspects of it.

Anomaly-based database intrusion detection systems are typically deployed to detect malicious accesses to the database by insiders where an insider is a person that belongs to an organization and is authorized to access a range of data and services. In literature, there are several anomaly-based database intrusion detection systems (for ease of exposition we refer to them as security-anomaly detection systems) that model normative behaviour of a user/role by considering queries made by that user/role to the database and subsequently normative profile is constructed using these model [21, 10, 9, 18, 5, 11, 14].

2.2. Adopted Privacy Model

Several forms of privacy have been formalized in the literature. The two mainstream definitions of privacy are k-anonymity [23] and differential privacy [7]. k-anonymity can be considered among the first formal definitions of privacy and serves as the foundation for several privacy definitions that includes l-diversity [17],

t-closeness [16], (α, k)-anonymity [25]. These privacy models are described within the framework of relational databases. We adopted k-anonymity firstly because this is an exploratory study, therefore, using a well-understood privacy model like k-anonymity enables a better understanding of the subject being explored and helps to avoid underlying complexities associated with other more complex privacy definitions. Secondly, k-anonymity served as a foundation of many subsequent formal privacy definitions, which is a good indicator of the applicability of this study onto other privacy definitions.

In the context of k-anonymity, attributes are classified in the following non-exclusive categories, *Identifiers*, *Quasi-Identifiers*, and *Sensitive attributes*. The classification is typically performed based on the risk of record re-identification using these attributes and the sensitivity of the information these attributes convey. An identifier is defined as *"an attribute that refers to only a particular individual in the given population* $\mathcal{U}$*"*. An example of an identifier is the Personal Public Service Number (PPS Number) which can uniquely identify individuals in Ireland. Other examples include an individual's passport number, driving license number, and e-mail address. Quasi-identifiers by themselves do not uniquely identify individuals; however, when correlated with other available external data, an individual (or individuals) can be identified. A quasi-identifier is defined in [23, 22] as a *"set of non-sensitive attributes of a relation if linked with external data to then uniquely identify at least one individual in the population* $\mathcal{U}$*"*. Let the set of quasi-identifiers be denoted as QI where each quasi-identifier is denoted as q_i. An example of quasi-identifier is the set of attributes `Zipcode`, `Date of Birth`, and `Gender`. For instance, the set of attributes `Zipcode`, `Date of Birth`, and `Gender` was used to re-identify governor of Massachusetts in [23, 22]. The re-identification was performed by directly linking shared attributes in two datasets, i.e. voter rolls and insurance company datasets. It was reported that 87% of the US population could be identified by these three attributes [23, 22]. Sensitive attributes consist of sensitive person-specific information. This information includes salary, disability status, or disease.

k-anonymity is defined in [23] as follows, *"a relation* $\mathcal{T}$ *satisfies k-anonymity if and only if each tuple* $r_i[QI] \in \mathcal{T}$ *appears with at least k occurrence in* $\mathcal{T}$*"*. k-anonymity provides a degree of anonymity if the data for each person cannot be distinguished from k-1 individuals in a released dataset with respect to a set of quasi-identifiers. Given QI then two tuples r_i and r_j are quasi-identifier equivalent if $r_i[QI] = r_j[QI]$. The relation $\mathcal{T}$ can be divided into quasi-identifier equivalence classes. Let the set of all the equivalence classes in $\mathcal{T}$ be $\mathcal{E}$ where each equivalence class $e \in \mathcal{E}$ consists of all the rows that have the same values for each quasi-identifier. Another way to define k-anonymity is that a relation $\mathcal{T}$ satisfies k-anonymity if the minimum equivalence class size is at least k in $\mathcal{T}$. Originally, k-anonymity was proposed for a one-time release of data, meaning that the user is not enabled to query the DBMS interactively. Though considered to be among the first privacy definitions, k-anonymity, has been widely applied in many domains to preserve privacy for examples Location-based services [28, 26, 27, 24, 30], ride-hailing services [15], and webmail auditing [6]. k-anonymity has been used along with cryptographic hashing to develop a protocol that provides a degree of anonymity while checking for passwords in a compromised databases [1].

3. Privacy-anomaly Detection System

This section described the notion of privacy-anomaly detection system. At design-level, the privacy-anomaly detection system has two phases, similar to conventional anomaly detection systems, that are a training (learning) phase and a detection phase. The normative k-anonymity based privacy-profile is mined in the training phase while in the detection phase a run-time privacy-profile is constructed and compared against the normative privacy-profile. In the next section, we also describe the structure of k-Anonymity based privacy-profile.

3.1. A k-Anonymity based Privacy-profile

In the proposed model k-anonymity is used to specify a privacy limit $[\![k\ ,\ q]\!]$, whereby k individual must share the same quasi identifier q values in the result of a query. Intuitively, this means for that particular response, for a sufficient value of k, an adversary can only narrow down to k individuals. In the case where an adversary has a secondary dataset with overlapping quasi-identifier values, then the query response can be

Table 1: A fragment of relation `temp_table`.

age	zipcode	city	gender	salary
>55	989234	Paris	Male	60K
>55	989234	Paris	Male	92K
>55	989234	Paris	Male	77K
>45	839523	Paris	Male	50K
>35	839777	London	Male	60K
>35	839777	London	Male	63K
>35	839777	London	Male	85K
>35	839777	London	Male	70K
>35	839777	London	Male	60K
>50	839567	Paris	Female	72K
>50	839567	London	Female	62K
>50	839567	Paris	Female	92K
>50	839567	London	Female	77K
>50	839567	Paris	Female	68K

linked to k different individuals, therefore minimizing the risk of re-identification. In the model the privacy-profile is defined as a set of privacy limits. In terms of privacy, each privacy limit means that in a particular instance of a query response an adversary won't be able to distinguish an individual's quasi-identifier values from k individuals for the set of quasi-identifiers that appeared in the query response. Consider a relation `temp_table`, as shown in Table 1, having several attributes including a sensitive attribute `salary`, and quasi-identifiers `age`, `gender`, `zipcode`, and `city`. For ease of exposition we assume the values for attribute `age` are aggregated into age ranges, for instance, all the values for attribute `age` above 55 are represented as >55. Given a mined privacy limit $[\![3, \{\texttt{age}, \texttt{zipcode}\}]\!]$, in privacy-profile, then the response to the analyst query `SELECT age, zipcode FROM temp_table WHERE gender = 'Male' AND city = 'Paris' AND age > 55;` as shown in Table 2 is not anomalous since the value of k for the the quasi-identifiers {`age`, `zipcode`} in the response is greater than 3.

Table 2: A relation $\mathcal{T}_{R1}$ resulting from the query `SELECT age, zipcode FROM temp_table WHERE gender = 'Male';`.

age	zipcode	salary
>55	989234	60K
>55	989234	92K
>55	989234	77K

Table 3: A relation $\mathcal{T}_{R2}$ resulting from the query `SELECT age, zipcode, city FROM temp_table WHERE gender = 'male';`.

age	zipcode	county	salary
>55	839523	Paris	60K
>55	839523	Paris	92K
>55	839523	Paris	77K
>45	839523	Paris	50K
>35	839777	London	60K
>35	839777	London	63K
>35	839777	London	85K
>35	839777	London	70K
>35	839777	London	60K

Table 4: A relation $\mathcal{T}_{R3}$ resulting from the query `SELECT age, zipcode FROM temp_table WHERE gender = 'female';`.

age	zipcode	salary
>50	839567	72K
>50	839567	62K
>50	839567	92K
>50	839567	77K
>50	839567	68K

3.2. Mining k-anonymity based Profiles for PAD

The instances of the privacy model are mined from audit logs in order to generate privacy-profiles. We refer to a privacy-profile that is mined from past logs in the learning phase as a normative privacy-profile. The idea is to learn the k values for sets of quasi-identifier(s) by mining past audit logs and interpret those mined 'privacy limits' as 'normal'.

Given an audit log L^*, consisting of query responses, $Pri(L^*)$ gives a privacy-profile consisting of privacy limits mined from log L^*, where $q \in QI$ represent a set of quasi-identifier. A normative privacy-profile is generated from an anomaly-free past log L^*_{norm} and is denoted by $Pri(L^*_{norm}) = \{ [\![k_1, q_1]\!], [\![k_2, q_2]\!], \ldots, [\![k_m, q_m]\!] \}$. For example, consider the relation $\mathcal{T}_{R2}$ shown in Table 3, the mined value of k for the set of quasi-identifiers

{age, zipcode, city} is 4, that is, $[\![4, \{age, zipcode, city\}]\!] \in Pri(L^*_{norm})$. In essence, we are constructing privacy limit (L^*, q) which returns k as a limit to the privacy in the table for a given q. The normative privacy-profile is effectively a set of these privacy limits mined against the logs for a given set of quasi-identifiers. Intuitively, the tuples in the normative privacy-profile shows to what extent one narrows down to individuals records in normative settings.

3.3. Detecting Privacy-anomalies

The detection phase, in terms of privacy, checks if an adversary is able to narrow down to fewer than k individuals for a given set of quasi-identifiers in the normative profile. In the instance, where the adversary is able to narrow down to fewer than specified k individuals for a given set of quasi-identifier then this instance is labelled as a privacy-anomaly and poses higher risk of re-identification relative to normal. During the detection phase, the run-time profile $Pri(L^*_{run})$ constructed given a run-time log L^*_{run}. $Pri(L^*_{run})$ is the constructed run-time profile. Given privacy limits $[\![k_i, q_i]\!]$ and $[\![k_j, q_j]\!]$ then $[\![k_i, q_i]\!]$ subsumes $[\![k_j, q_j]\!]$ (denoted $[\![k_i, q_i]\!] \preceq [\![k_j, q_j]\!]$) if imposing privacy limit $[\![k_j, q_j]\!]$ instead of $[\![k_i, q_i]\!]$ leads to no additional loss of privacy. Formally,

$$[\![k_i, q_i]\!] \preceq [\![k_j, q_j]\!] \equiv q_i \subseteq q_j \wedge k_j \geq k_i$$

In the case where $[\![k_i, q_i]\!] \in Pri(L^*_{norm})$ and $[\![k_j, q_j]\!] \in Pri(L^*_{run})$ then $[\![k_i, q_i]\!] \preceq [\![k_j, q_j]\!]$ means that $[\![k_j, q_j]\!]$ can be safely replaced by $[\![k_i, q_i]\!]$ without any loss of privacy. If a privacy limit subsumes another intuitively it means if the subsumed privacy limit is replaced by the one that subsumes it then there is no loss of privacy.

Consider the response of a query at run-time shown in Table 4, and that there exists a privacy limit $[\![3, \{age, zipcode\}]\!]$ in $Pri(L^*_{norm})$. The mined value k of the set of quasi-identifier $\{age, zipcode\}$ is greater than 3 therefore this privacy limit $[\![5, \{age, zipcode\}]\!]$ in $Pri(L^*_{run})$ is considered to be subsumed by the privacy limit $[\![3, \{age, zipcode\}]\!]$ in $Pri(L^*_{norm})$. In terms of privacy, it means given that this instance of query response an adversary can narrow down so many individuals as one normally is able to for a given set of quasi-identifiers.

4. Correlation between Security and Privacy-anomalies

This section aims to discover whether anomalies labelled by a privacy-anomaly detection system presented in Section 3.2 are labelled as anomalies by the conventional anomaly detection systems. For the sake of clarity, we call the conventional anomaly detection systems as a security-anomaly detection system. One such security-anomaly detection system is proposed in [10].

4.1. Security-anomaly Detection System based on n-gram

The security-anomaly detection system in [10], models normative behaviours using n-grams of normal query patterns extracted from the audit log of SQL queries of an application system. SQL queries were transformed into an abstract representation. Subsequently, a normative profile was constructed that consisted of sets of n-grams of SQL query abstractions. For a given sequence L of SQL queries, $abs(L)$ represents the abstraction of SQL queries in L and $ngram(abs(L),n)$ is the set of all sub-sequences of size n that appear in $abs(L)$. Lets say $abs(L) = abs(Q_1), abs(Q_2), abs(Q_3), abs(Q_4)$ then a 2-gram model for $abs(L)$ will be $\{\langle abs(Q_1), abs(Q_2)\rangle, \langle abs(Q_2), abs(Q_3)\rangle, \langle abs(Q_3), abs(Q_4)\rangle\}$.

4.2. Discovering Correlations

This section explores whether privacy-anomalies (as identified by the model in Section 3.2) are also identified as security-anomalies by a security-anomaly detection system in [10]). The security-anomaly detection system in [10] relies on n-grams to construct profiles of querying behaviours using audit logs of SQL queries.

The system in [10] effectively detects malicious accesses by insider to a database management system. A query generator was designed that had defined a set of SQL query templates and the underlying database was populated with a synthetic (hospital) dataset. Query templates were designed to be executed on the hospital dataset and mimicked the health-care analytics scenario. A fragment of the dataset is shown in Table 5.

Logs were generated for construction of a normative profile and another for the construction of a run-time profile. The training logs (anomaly-free) for the n-gram based approach are denoted by L^{hosp}_{norm}, while the anomalous run-time logs for the hospital datasets are denoted by L^{hosp}_{run}.

To construct normative and run-time profiles using the n-gram model, selection of an appropriate value of the size of n-gram was desirable for the hospital dataset. To select an appropriate size of an n-gram in this scenario, test logs L^{hosp}_{test1} and L^{hosp}_{test2} were generated in a safe environment (anomaly-free). N-gram profiles were constructed with varying n-gram size, that are, $ngram(L^{hosp}_{test1}, n)$ and $ngram(L^{hosp}_{test2}, n)$, and generated profiles were compared. Figure 1 depicts the number of n-gram mismatches arising when comparing the normal test $ngram(L^{hosp}_{test1}, n)$ and $ngram(L^{hosp}_{test2}, n)$, for different values of n. From the experiments, the n-gram of the size of 4 ($n = 4$) was considered optimal as it resulted in an acceptable number of mismatches.

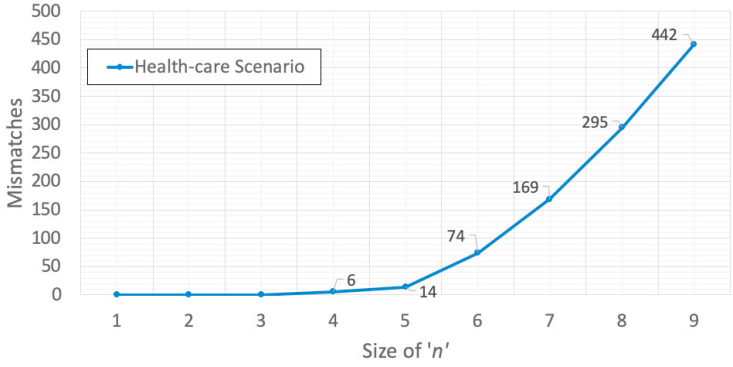

Figure 1: The figure shows the number of mismatches between $ngram(L^{hosp}_{test1}, n)$ and $ngram(L^{hosp}_{test2}, n)$ for different values of n.

Once the value of n was decided upon, the normative and run-time profiles were constructed for the experiments. Given the training logs L^{hosp}_{norm} and L^{hosp}_{run} n-gram profiles were constructed such that $ngram(L^{hosp}_{norm}, 4)$ and $ngram(L^{hosp}_{run}, 4)$, and subsequently the normative and runtime profiles were compared.

The same queries in logs L^{hosp}_{norm} and L^{hosp}_{run} were executed in the presence of the privacy-anomaly detection system (described in Section 3) resulting in logs of query responses L^{hosp*}_{norm} and L^{hosp*}_{run}. Subsequently, a normative privacy-profile $Pri(L^{hosp*}_{norm})$ and a run-time $Pri(L^{hosp*}_{run})$ profiles were constructed and compared.

The attribute `patient_ID` and `e-mail_ID` were considered as a unique identifier, the attribute `diagnosis` was considered as a sensitive attribute while the rest of the attributes including `first_name`, `last_name`, `status`, `dob`, `gender`, `city`, and `marital_status` were considered as quasi-identifiers. For the experimentation, two categories of privacy-anomalies were injected as described in Table 6. Using this anomaly-containing run-time log, from 15 privacy-anomalies 13 were detected by the n-gram based security-anomaly detection system proposed in [10] and the privacy-anomaly detection system proposed in this paper.

4.3. Detected Privacy-anomalies

The n-gram based security-anomaly detection system detected all those privacy-anomalies that were generated by injecting one more attribute into the relation. The privacy-anomalies injected by adding one more attribute were identified as privacy-anomalies by both systems. The reason that they were identified was because there were no n-gram that contained a reference to new attribute in its query abstraction.

One of the detected privacy-anomalies corresponds to the query: `SELECT diagnoses, dob, city, country FROM hospitalDB WHERE dob = '1981' AND city = 'London';`

Table 5: A fragment of hospital dataset. The strike-through attribute values represents a deleted row.

dob	city	gender	diagnoses	country	...
1981	London	Male	Flu	UK	...
1981	London	Male	Flu	UK	...
1981	London	Male	Diarrhoea	US	...
1920	Paris	Male	Heart Disease	UK	...
1981	Berlin	Female	Acne	UK	...
1984	Berlin	Male	Flu	Australia	...
1984	Berlin	Male	Diabetes	UK	...
1984	Berlin	Male	Hypertension	UK	...
1984	Berlin	Male	Leg Fracture	Portugal	...
...	...	...	...	...	...
...	...	...	...	...	...
~~1981~~	~~London~~	~~Male~~	~~Flu~~	~~US~~	...

Table 6: Description of Privacy-anomalies injected.

Description of privacy-anomalies	Number of anomalies injected
Addition of one or more attributes to the base relation shown in Table 5. For instance, a new attribute, like `country`, was inserted in the relation and queries were made to retrieve this attribute values.	5
Update or Deletion of records from relation shown in Table 5	10

Table 7: Response to a undetected privacy-anomalous query.

dob	city	diagnoses
1920	Paris	Heart Disease

The normative privacy-profile contains no privacy limit reference to the new (or combination of new) attribute.

4.4. Undetected Privacy-anomalies

A privacy-anomaly undetected by the n-gram based approach but detected by the privacy model is: `SELECT dob, city, diagnoses FROM hospitalDB WHERE dob = '1920' AND city = 'Paris' ;`

The query returns a relation with one record as shown in Table 7. It is identified as a privacy-anomaly by the privacy model for the reason being that the specified value of k for the specified set of quasi-identifier meant that an adversary was able to single out an individual. This anomaly is undetected by n-gram based security-anomaly detection approach because there was an n-gram in normative profile contained a reference to this query abstraction.

In the examples above, the privacy-anomalies illustrated are based on a single query rather than a query sequence.

4.5. Identifying Appropriate Privacy Limits

In order to find the optimal values of k, in the mining process, in theory, all the combinations of quasi-identifiers need to be considered. This, in essence, is a combinatorial explosion, especially in the case of a large number of quasi-identifiers. Additionally, one may discover either very large or very small values of k in practice for certain combinations of quasi-identifiers. Therefore, in order to discover reasonable values of k, one may define a range while mining the values of k such that the values falling within the range and their corresponding combinations of quasi-identifiers are considered for privacy-profiles.

5. Conclusions

The paper proposed the notion of privacy-anomaly detection and described a naïve instantiation based on well-known privacy model, that is, k-anonymity. The idea is to model normative privacy by mining privacy

limits from logs for past interaction and construct normative privacy-profiles in the training phase. While in the detection phase, a run-time privacy-profile is constructed and checked against normative privacy limits in normative privacy-profile. The deviations between the normative privacy profile and run-time privacy profile are labelled as privacy-anomalies. As this is an exploratory study, therefore, *k*-anonymity is used as it served as a foundation of many subsequent formal privacy definitions, which is a good indicator of the applicability of this study onto other privacy definitions. Furthermore, the naïve instantiation of privacy-anomaly detection system was demonstrated in this paper over a synthetic dataset. The paper also considered the question of whether there is a correlation between security-anomalies (access control anomalies) and the privacy-anomalies. For this study, a conventional security-anomaly that detects database access by malicious insiders (employees of an organisation) was adopted. The security-anomaly detection system used n-grams to model query behaviours. It was discovered that conventional security-anomaly detection system labelled some of the privacy-anomalies while some of the privacy-anomalies went undetected.

In future, we plan to further the work by investigating advanced instantiation of a privacy-anomaly detection system based on the composition of several privacy models. A potential instantiation of privacy-anomaly detection system is to mine normative privacy-profiles in-terms of identification capabilities of SQL queries [12] made to DBMS.

Acknowledgements

This material is based upon the work supported by the Science Foundation Ireland under Grant number 12/RC/2289 which is co-funded under the European Regional Development Fund.

References

[1] Ali, J., 2017. Mechanism for the prevention of password reuse through anonymized hashes. PeerJ PrePrints 5, e3322. URL: https://doi.org/10.7287/peerj.preprints.3322v1, doi:10.7287/peerj.preprints.3322v1.

[2] Barbaro, M., Jr., T.Z., . A face is exposed for aol searcher no. 4417749. URL: http://www.nytimes.com/2006/08/09/technology/09aol.html?mcubz=2. the New York Times. Online at: http://www.nytimes.com/2006/08/09/technology/09aol.html?mcubz=2.

[3] Bhuyan, M.H., Bhattacharyya, D.K., Kalita, J.K., 2014. Network anomaly detection: Methods, systems and tools. IEEE Communications Surveys Tutorials 16, 303–336. doi:10.1109/SURV.2013.052213.00046.

[4] Chandola, V., Banerjee, A., Kumar, V., 2009. Anomaly detection: A survey. ACM Comput. Surv. 41, 15:1–15:58. URL: http://doi.acm.org/10.1145/1541880.1541882, doi:10.1145/1541880.1541882.

[5] Costante, E., den Hartog, J., Petković, M., Etalle, S., Pechenizkiy, M., 2017. A white-box anomaly-based framework for database leakage detection. J. Inf. Secur. Appl. 32, 27–46. URL: https://doi.org/10.1016/j.jisa.2016.10.001, doi:10.1016/j.jisa.2016.10.001.

[6] Di Castro, D., Lewin-Eytan, L., Maarek, Y., Wolff, R., Zohar, E., 2016. Enforcing k-anonymity in web mail auditing, in: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, ACM, New York, NY, USA. pp. 327–336. URL: http://doi.acm.org/10.1145/2835776.2835803, doi:10.1145/2835776.2835803.

[7] Dwork, C., 2008. Differential privacy: A survey of results, in: Agrawal, M., Du, D., Duan, Z., Li, A. (Eds.), Theory and Applications of Models of Computation, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 1–19.

[8] Hern, A., 2014. New york taxi details can be extracted from anonymised data, researchers say. URL: https://www.theguardian.com/technology/2014/jun/27/new-york-taxi-details-anonymised-data-researchers-warn. the Guardian. Online at: https://www.theguardian.com/technology/2014/jun/27/new-york-taxi-details-anonymised-data-researchers-warn.

[9] Hussain, S.R., Sallam, A.M., Bertino, E., 2015. Detanom: Detecting anomalous database transactions by insiders, in: Proceedings of the 5th ACM Conference on Data and Application Security and Privacy, ACM, New York, NY, USA. pp. 25–35. URL: http://doi.acm.org/10.1145/2699026.2699111, doi:10.1145/2699026.2699111.

[10] Khan, M.I., Foley, S.N., 2016. Detecting anomalous behavior in DBMS logs, in: Cuppens, F., Cuppens, N., Lanet, J., Legay, A. (Eds.), Risks and Security of Internet and Systems - 11th International Conference, CRiSIS 2016, Roscoff, France, September 5-7, 2016, Revised Selected Papers, Springer. pp. 147–152. URL: https://doi.org/10.1007/978-3-319-54876-0_12, doi:10.1007/978-3-319-54876-0_12.

[11] Khan, M.I., Foley, S.N., O'Sullivan, B., 2018a. Dbms log analytics for detecting insider threats in contemporary organizations, in: Abassi, R., Douss, A.B.C. (Eds.), Security Frameworks in Contemporary Electronic Government, IGI Global. pp. 207–234.

[12] Khan, M.I., Foley, S.N., O'Sullivan, B., 2019. Computing the identification capability of sql queries for privacy comparison, in: Proceedings of the ACM International Workshop on Security and Privacy Analytics, Association for Computing Machinery, New York, NY, USA. pp. 47 – 52. URL: https://doi.org/10.1145/3309182.3309188, doi:10.1145/3309182.3309188.

[13] Khan, M.I., O'Sullivan, B., Foley, S.N., 2018b. A semantic approach to frequency based anomaly detection of insider access in database management systems, in: Cuppens, N., Cuppens, F., Lanet, J.L., Legay, A., Garcia-Alfaro, J. (Eds.), Risks and Security of Internet and Systems, Springer International Publishing, Cham. pp. 18–28.
[14] Khan, M.I., O'Sullivan, B., Foley, S.N., 2018. Towards modelling insiders behaviour as rare behaviour to detect malicious rdbms access, in: 2018 IEEE International Conference on Big Data (Big Data), pp. 3094–3099.
[15] Khazbak, Y., Fan, J., Zhu, S., Cao, G., 2018. Preserving location privacy in ride-hailing service, in: 2018 IEEE Conference on Communications and Network Security (CNS), pp. 1–9. doi:10.1109/CNS.2018.8433221.
[16] Li, N., Li, T., Venkatasubramanian, S., 2007. t-closeness: Privacy beyond k-anonymity and l-diversity, in: 2007 IEEE 23rd International Conference on Data Engineering, pp. 106–115. doi:10.1109/ICDE.2007.367856.
[17] Machanavajjhala, A., Kifer, D., Gehrke, J., Venkitasubramaniam, M., 2007. L-diversity: Privacy beyond k-anonymity. ACM Trans. Knowl. Discov. Data 1. URL: http://doi.acm.org/10.1145/1217299.1217302, doi:10.1145/1217299.1217302.
[18] Mathew, S., Petropoulos, M., Ngo, H.Q., Upadhyaya, S., 2010. A data-centric approach to insider attack detection in database systems, in: Proceedings of the 13th International Conference on Recent Advances in Intrusion Detection, Springer-Verlag, Berlin, Heidelberg. pp. 382–401. URL: http://dl.acm.org/citation.cfm?id=1894166.1894192.
[19] Narayanan, A., Shmatikov, V., 2008. Robust de-anonymization of large sparse datasets, in: 2008 IEEE Symposium on Security and Privacy (SP' 08), IEEE Computer Society, Los Alamitos, CA, USA. pp. 111–125. URL: https://doi.ieeecomputersociety.org/10.1109/SP.2008.33, doi:10.1109/SP.2008.33.
[20] Pieczul, O., Foley, S.N., 2016. Runtime Detection of Zero-Day Vulnerability Exploits in Contemporary Software Systems. Springer International Publishing, Cham. pp. 347–363.
[21] Sallam, A., Fadolalkarim, D., Bertino, E., Xiao, Q., 2016. Data and syntax centric anomaly detection for relational databases. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 6, 231–239. URL: http://dx.doi.org/10.1002/widm.1195, doi:10.1002/widm.1195.
[22] Sweeney, L., 2000. Simple Demographics Often Identify People Uniquely. Working paper. Working paper. Online at: http://dataprivacylab.org/projects/identifiability/.
[23] Sweeney, L., 2002. k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10, 557–570. URL: https://doi.org/10.1142/S0218488502001648, doi:10.1142/S0218488502001648.
[24] Wang, Y., Cai, Z., Chi, Z., Tong, X., Li, L., 2017. A differentially k-anonymity-based location privacy-preserving for mobile crowdsourcing systems, in: Bie, R., Sun, Y., Yu, J. (Eds.), 2017 International Conference on Identification, Information and Knowledge in the Internet of Things, IIKI 2017, Shandong, China, October 19-21, 2017, Elsevier. pp. 28–34. URL: https://doi.org/10.1016/j.procs.2018.03.040, doi:10.1016/j.procs.2018.03.040.
[25] Wong, R.C.W., Li, J., Fu, A.W.C., Wang, K., 2006. (α, k)-anonymity: An enhanced k-anonymity model for privacy preserving data publishing, in: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, NY, USA. pp. 754–759. URL: http://doi.acm.org/10.1145/1150402.1150499, doi:10.1145/1150402.1150499.
[26] Ye, Y.M., Pan, C.C., Yang, G.K., 2016. An improved location-based service authentication algorithm with personalized k-anonymity, in: Sun, J., Liu, J., Fan, S., Wang, F. (Eds.), China Satellite Navigation Conference (CSNC) 2016 Proceedings: Volume I, Springer Singapore, Singapore. pp. 257–266.
[27] Zhang, Y., Tong, W., Zhong, S., 2016. On designing satisfaction-ratio-aware truthful incentive mechanisms for k-anonymity location privacy. IEEE Transactions on Information Forensics and Security 11, 2528–2541. doi:10.1109/TIFS.2016.2587241.
[28] Zhao, P., Li, J., Zeng, F., Xiao, F., Wang, C., Jiang, H., 2018. Illia: Enabling k-anonymity-based privacy preserving against location injection attacks in continuous lbs queries. IEEE Internet of Things Journal 5, 1033–1042. doi:10.1109/JIOT.2018.2799545.
[29] Zhao, Q., Zhang, Y., Shi, Y., Li, J., 2020. Analyzing and visualizing anomalies and events in time series of network traffic, in: Boonyopakorn, P., Meesad, P., Sodsee, S., Unger, H. (Eds.), Recent Advances in Information and Communication Technology 2019, Springer International Publishing, Cham. pp. 15–25.
[30] Zhong, S., Zhong, H., Huang, X., Yang, P., Shi, J., Xie, L., Wang, K., 2019. Connecting Things to Things in Physical-World: Security and Privacy Issues in Vehicular Ad-hoc Networks. Springer International Publishing, Cham. pp. 101–134. URL: https://doi.org/10.1007/978-3-030-01150-5_5, doi:10.1007/978-3-030-01150-5_5.